*The Geographical Journal*, Vol. **154**, No.1, March 1988, pp. 56–68

# THE DOMESDAY MACHINE: A NATIONWIDE GEOGRAPHICAL INFORMATION SYSTEM

DAVID RHIND, PETER ARMSTRONG AND STAN OPENSHAW

This paper describes selected aspects of the Domesday project led by the BBC and intended to create 'an exhibition of Britain' in the 1980s. Among other things, this resulted in what is perhaps the first example of a second-generation Geographical Information System (GIS). Based upon a micro-computer linked to a new Laser Vision-Read Only Memory (LV-ROM), this holds 54 000 images (maps, photos, satellite images, etc.), 300 megabytes of digital data and millions of words of text per side of video disk; the different types of data are cross-referenced by geographical position or by theme. Access to the data is by pointing at maps, by specification of place name or geographical coordinates, or through use of a thesaurus: the source and storage form of the data is transparent to the user. Included in the initial disks are 21 000 files of spatial data showing national coverage down, in some cases, to 1 square kilometre resolution. Data sets stored include geology, soils, geochemistry, population, employment and unemployment, agricultural production and land use/land cover. Though the normal purchase price of the data held would be over £250 000, the price charged to schools for the complete system—hardware, software and data—at the launch date in November 1986 was £3000. All this has significant implications for the spread of use of geographical data bases and GIS technology; it also has major implications for the teaching of Geography at different levels.

KEY WORDS: Geographical Information Systems, cartography, statistics, computer, United Kingdom.

IN 1086, THE NOW WELL-KNOWN survey of much of England was carried out at the behest of William the Conqueror. Nine hundred years later, a collaborative project, led by the BBC, has repeated the exercise, extending its scope to cover all of Britain and including an enormously increased range of information. The results of this £3 million project include:

- a micro-computer system, part of which is a Geographical Information System (GIS);
- a new video disk player, capable of storing and overlaying information held in both analogue and digital form;
- two video disks containing 30 million words, 21 000 files of spatial (or mappable) digital data, 24 000 Ordnance Survey topographic maps, statistical tabulations and time series, picture libraries and TV film clips (see Fig. 1).

As a consequence, the first ever country-wide computerized geographical data base has been produced and is readily available. No such computer system or data exists anywhere else in the world.

The project effectively began in December 1984, the launch of the final product being 23 months later. In that time, the following were carried out: the development of data storage and access concepts; the design and construction of the hardware and of the software; negotiations to obtain data sets from government, private sector, academic and other agencies; the organization of 14 000 schools to collect certain types of information; the validation of the data; and the construction of documentation. Details of the history of the project and of the organization of the various teams working in different parts of Britain are not the concern of this paper (see Goddard and Armstrong, 1986). Equally, the voluminous non geographical data, both analogue and

→ Professor David Rhind is Professor of Geography, Birkbeck College, University of London, Malet Street WC1 7HX. Peter Armstrong, BBC Television, devised and directed the project for the BBC. Dr Stan Openshaw of the University of Newcastle's Centre for Urban and Regional Development Studies was leader of one of the four supporting teams based in universities. This paper follows a presentation given in the Society's House in February 1987.

0016-7398/88/0001-0056/$0.20/0

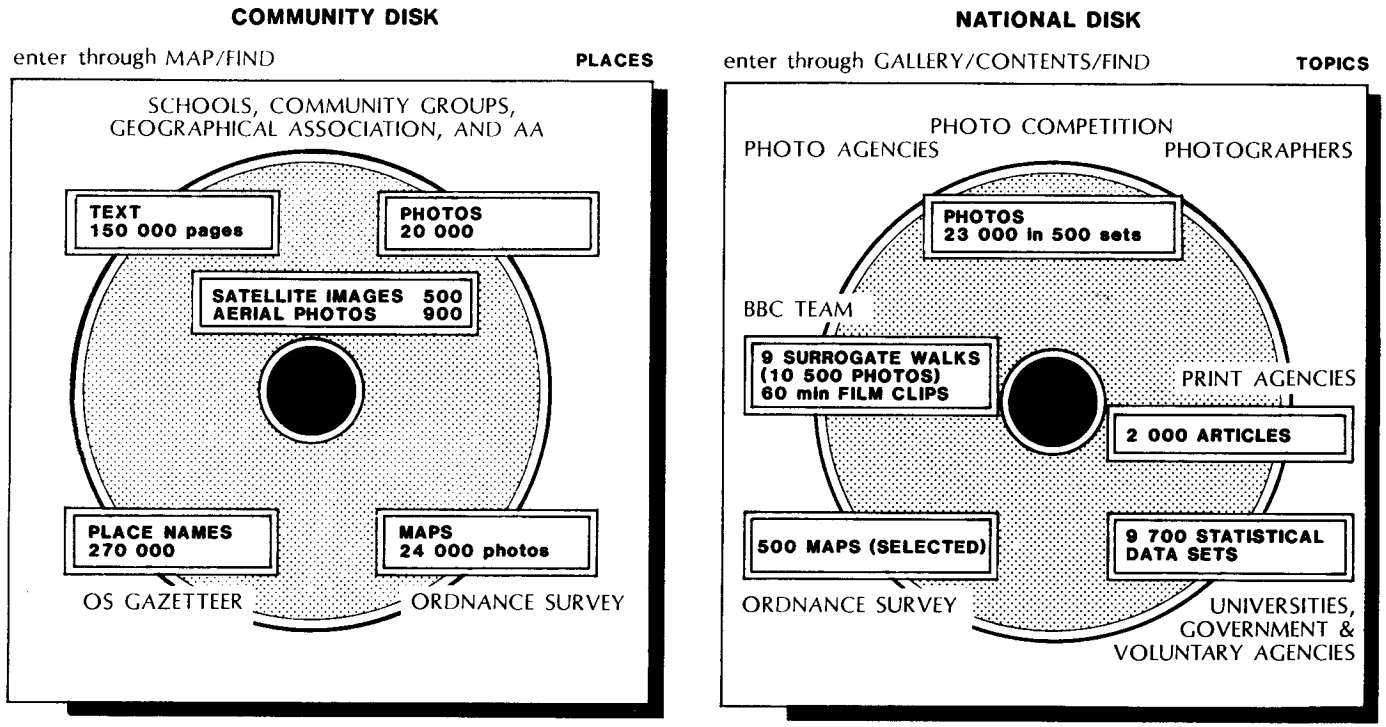



*Fig. 1. General overview of the types of information held on the two Domesday disks, the data sources and the methods of accessing these data*

*Source:* BBC Enterprises

digital, are not relevant here although they include numerous libraries of photographs on topics as diverse as the Royal Family, British Design Council Award winners, ceramics, and public houses. Figure 1, however, summarizes some of the sources of the information and indicates the complications of project management. Because of their intrinsically geographical nature, this paper concentrates on those spatially-related facilities which the Domesday machine makes available.

The Domesday machine can be considered the first example of a second generation GIS. The justification of this claim is as follows:

(i) it handles data in both analogue and digital form and permits graphic overlay of one on the other, plus some digital operations in relation to analogue maps. (In this case analogue refers to photographic copies of paper documents such as maps.)
(ii) it comes complete with its own data base, currently totalling in excess of 300 megabytes and covering a vast range of environmental, demographic, socio-economic and other variables.
(iii) it provides high response rate interactive colour graphics.
(iv) it implements the cross-linkage of maps, air photographs, colour slides, moving pictures, text (held digitally)and digital statistical data. Thus, a user interested in one geographical area can move from one type of information to another, virtually instantaneously.
(v) it is extremely easy to use. Successful demonstrations to the Prime Minister were given by 11-year old school children who had only two hours practice on the system.
(vi) it is very inexpensive, being designed (by previous GIS standards) for a mass market. Thus schools can purchase the entire hardware, software and data for only £3000 whilst other purchasers pay £4600 including VAT. This should be contrasted with the normal purchase price of only those digital data sets on the Domesday disks which are readily available; their cost would exceed £250 000.
(vii) despite (v) and (vi), it includes several desirable capabilities which are not commonplace, yet are of vital importance to all studying geographical data; it permits the user, for instance, to study the effects of changing the data resolution along various scale hierarchies (e.g. electoral wards to administrative districts to counties, to regions, of 1, 2, 3 . . . 10 km squares).

This paper describes the computer system and the data base in more detail before concluding with a consideration of the likely social, educational, and administrative effects of the advent of the Domesday machine.

*The computer hardware and software*

The initial release of the system consists of:

- a BBC 128k Master Series micro-computer, including floppy disk drives and tracker-ball;
- a new Philips Laservision (LV) 12 in Read Only Memory (ROM);
- a high resolution colour monitor;
- retrieval and analysis software;
- two Domesday disks, the national and the local (or 'Community') disks.

Openshaw, Wymer and Charlton (1986) have described the basis of the system in some detail; Plate I illustrates the components. A rather more powerful micro (the RML Nimbus) is also now available to drive the system whilst other computers, notably an Atari and IBM PC-compatible versions are scheduled to be available by the end of 1987.

In operation, extensive use is made of the tracker-ball to access the data, by pointing at items in menus, at positions on maps or on keys in statistical displays. The 'default access' to the National Disk, for instance, is by 'navigating' a picture gallery (Plate VII(c)), each picture representing a topic which may be pursued; alternatively, the user may 'walk' out of any one of the doors into different types of environment. In addition,

keywords or keyword strings, place names, National Grid Reference coordinates and other items may be entered in the normal way via the keyboard.

The software is written in BCPL and was produced by Logica Limited under contract to the BBC. The data structure utilized for storing the spatial ('mappable') data files was originally devised by Stan Openshaw and colleagues and tested on a VAX computer. In essence, all data on the system are held in raster (or grid cell) form because of memory limitations in the initial micro-computer. Thus attribute data for administrative areas are stored as fixed length lists; the length of the list reflects the number of areal units. The vector boundaries of the areas are held in highly compacted form, with pointers between the two data sets. Default values for class intervals and various other characteristics of each of the 21 000 data files of this type were computed and stored at the time the data were being created.

*The data base*

Four main contractors were charged with obtaining and/or reorganizing data for the Domesday disks. These were the Birkbeck College Department of Geography, the Centre for Urban and Regional Development Studies at the University of Newcastle, the Economic and Social Research Council Data Archive at Essex University and the Institute of Terrestrial Ecology, Bangor. The different teams had usually disparate but sometimes overlapping responsibilities: as a consequence, collaboration was essential and constant use was made of the UK Joint Academic Network (JANET) (Wells, 1984) for electronic mail and for the transfer of certain data sets between the collaborating groups.

The initial Domesday 'package' includes two video disks, of which the contents and means of access are summarized in Figure 1. Selected aspects of each disk will now be considered.

*The local or Community disk* This is a 'people's data base' on Britain, in so far as much of it was compiled by nearly one million individuals—mostly schoolchildren—and represents the aggregate of their views on small areas of the country. The bulk of information on this disk consists of Ordnance Survey (OS) topographic maps, including complete coverage of the country based upon 1/50 000, 1/250 000 and smaller scale maps and 1/10 000 scale maps for 80 cities (see Plate II); larger scale maps, floor plans etc. of sites of special interest; Landsat Thematic Mapper and Multi-Spectral Scanner imagery; 30 million words of descriptive text; colour slides of locations throughout the country; and the OS gazetteer. All information on this disk, apart from the text, is held in analogue or picture form. The information is arranged hierarchically as shown in Table I; 'zooming in' and out between these levels and hence changing the level of geographical detail, scrolling across country and changing from text to pictures to topographic maps for the same area is virtually instantaneous and is effected by use of the tracker-ball. Moreover, digital operations on the analogue maps, such as measuring length and area, are possible (see Plate IV).

The level 0 and 1 essays were written by academic geographers, whilst those at level 2 were written by other professionals, ranging from school teachers to university professors. The level 3 text is extremely heterogeneous, being written in many cases by school children and is available for about nine thousand 4 × 3 km areas in Britain. In itself, it comprises an unsurpassed social, as well as geographical, record of its times; much of the text, for instance, was written during the miner's strike of 1984–85 and geographical variations in attitudes to it are very evident.

Access to the information held on the Community disk is by two mechanisms. The first is by pointing at locations on a map, then 'zooming in' to obtain more detail, or selecting via the menu other information in relation to that place. The alternative is to use the 'Find' command and specify a place name; the computer then searches the 270 000 names in the gazetteer, containing all places on the OS 1/50 000 scale maps, and plots the appropriate map, highlighting the 1 km square in which the name is located.

*The national disk* The data on the national disk are of three main types; a set of picture libraries showing many aspects of British life in the 1980s; several thousand

TABLE I

| *Level* | *Area name/size* | *Data held* |
|---|---|---|
| 0 | UK | Essay: Landsat MSS mosaic |
| 1 | Regions (N. and S. Britain, N. Ireland, Isle of Man, Orkney and Shetlands, Channel Islands) | Essays: Landsat MSS mosaics |
| 2 | 40 × 30 km blocks (covering *c.* 70% of the UK) | Essays, Landsat TM true colour and false colour images and up to 5 air photographs |
| 3 | 4 × 3 km blocks (covering *c.* 45% of the UK) | Essays, up to 4 colour slides |

*Key:* TM Thematic Mapper; MSS Multi-Spectral Scanner

cross-tabulations of statistical data derived from government series such as the Family Expenditure Survey; and also a variety of spatial data. Table II illustrates the highest resolution spatial data sets which pertain directly to the land and, less directly, to people. Numerous other data sets exist on the disk: some of the socio-economic data are described by Owen *et al.* (1986). In general, and as a matter of principle, the data are held and are available by whatever 'standard' geographical areas were used to report them by the original data compilers (such as the Department of Employment); these areas include grid squares, parliamentary constituencies, districts, counties, functional regions, Local Education Authority areas, television regions and Regional Health Authority areas. Thirty-three different sets of areas were used in total. Many of these data sets, however, are held at multiple levels of geographical aggregation.

It should be emphasized that the data listed in Table II are only the so-called 'spatial' data, i.e. those which can be mapped and (given suitable software) manipulated on a spatial basis. Many other environmental data sets exist on the disk which can be tabulated and most of these 'non spatial' data sets are crudely classified by geographical area; forestry data, for instance, are only available in 'non spatial' form but include details of the areal extents of different species of woodland trees for each county in Britain (see Plates VI and VII (a)).

In addition to the variables drawn from official data sources, a schools-based project led to the collection of nearly 70 items of data for each of over 100 000 1 km grid squares. Nearly half the schools in Britain took part in this data-gathering exercise in which primary, secondary and tertiary land use and land cover were recorded for each 1 km square, together with the number of occurrences of facilities such as banks, leisure centres and schools (Rhind and Mounsey, 1986). These data were checked on a spot check basis in Birkbeck College and also checked exhaustively by comparing each and every value given for each 1 km square with a 'feasibility value' nominated for each separate variable by experience.

Some aspects of the compilation of the national disk data were novel. Perhaps the best example of this is Green's creation of a 1 km grid data set of Population Census data for the whole of mainland Britain. He took the data for 125 000 small but irregularly-sized Enumeration Districts (EDs), spatially described only by a coordinate pair locating the centroid of the area; from this, he generated a Dirichlet tesselation, clipped this by superimposing a detailed coastline of the whole country, rasterised these Dirichlet tiles to 100 m resolution—producing, in effect, a 12 000 by 8000 matrix of ED names—recombined these into 1 km areas by allocating the appropriate fraction of each ED's population to the larger grid cell, restored the unpopulated areas known to exist by a thresholding process and ensured that county population totals matched published figures. The entire operation is described in more detail in Rhind and Mounsey (1986, p. 323) and is believed to be unprecedented on a countrywide scale. Whilst the results are not as good as those obtained by genuine grid square data (e.g. see CRU/OPCS/GRO(S), 1981) they are much more manageable (and hence readily

PLATE I

PLATE II

PLATE III

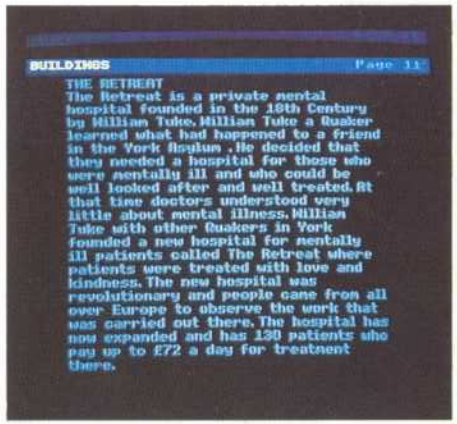



PLATE IV


PLATE I *The Domesday machine, showing the BBC micro-computer, the Philips LV-ROM, the display, a tracker-ball and one laser disk*

PLATE II *A sample of an OS 1/10 000 scale map photographed off the screen (Crown Copyright reserved)*

PLATE III *A sample of the 30 million words of text: the first page of a Level 3 essay written by schoolchildren*

PLATE IV *Measuring the distance between York Minster and York University on an OS 1/50 000 scale map; the 'path' may be made as complex as is required (Crown Copyright* reserved on the map)

PLATE V

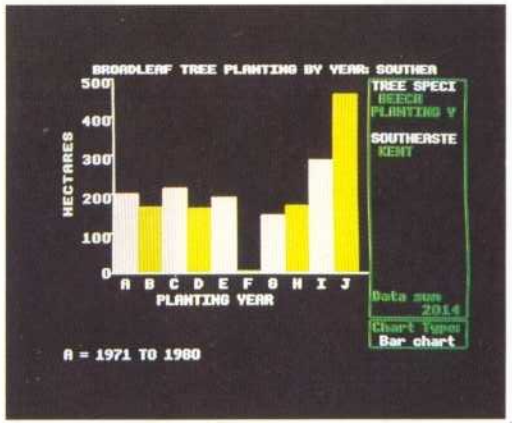



PLATE VI

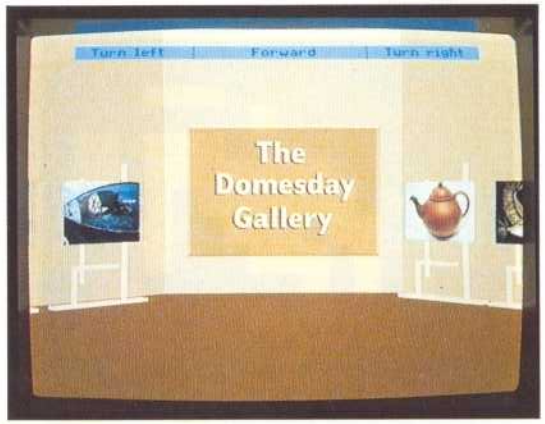
PLATE V *The distribution of three crops*

PLATE VI *A histogram of statistics of beech tree planting in Kent between 1971 and 1981*

a

b

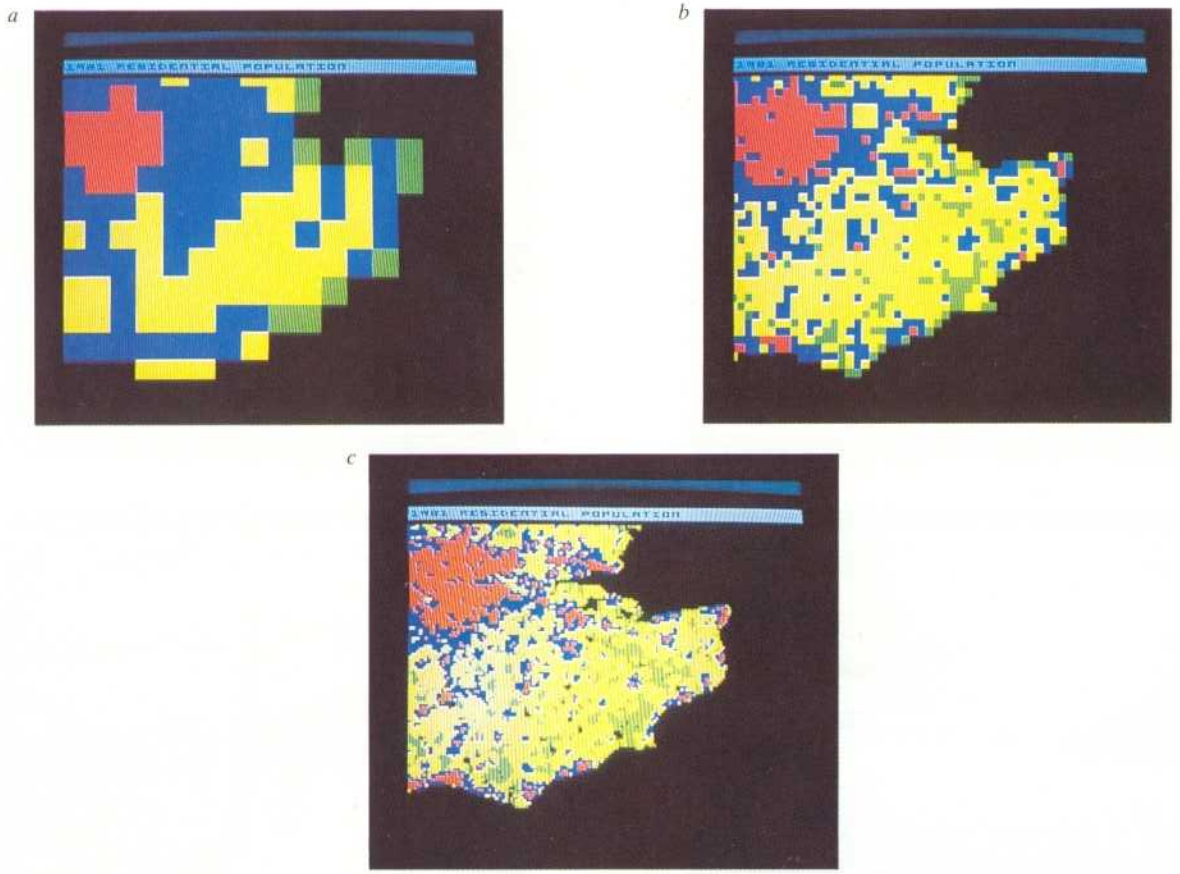
c




d


*(a) Photograph of a beech wood (b) Coastal geomorphology in Sussex*
*(c) The Domesday gallery—the default way into National Disk information*
*(d) Scene from a 'surrogate walk' around Brecon. The 'looking glass' indicates a feature on which detailed information is available*

PLATE VIII




*Population density in Kent and surrounding areas in 1981, shown at (a) 10 km (b) 3 km and (c) 1 km resolution. The differences in structure of the map are evident; similarly large differences obtained in statistical analysis of the data (the census data from which these data sets were derived are Crown Copyright)*

TABLE II

| *Data* | *Source* | *Resolution (n × n km)* | *Coverage* | *Location* |
|---|---|---|---|---|
| Place names | OS | 1 | UK | C |
| Physiography and topography | | | | |
| (maps) | OSGB + OSNI | varies with scale | GB + NI | C + N |
| (data) | ITE | 10 | UK | N |
| (text) | schools, etc. | various | 45–70% GB | C |
| Land use cover | schools, etc. | 1 | 45% GB | C |
| Incidence counts | schools, etc. | 1 | 50% GB | C |
| Solid geology | BGS | 1 | GB | N |
| Drift geology | ITE | 10 | UK | N |
| Soils | Soil Surveys of | | | |
| | England and Wales | 5 | E + W | N |
| | and of Scotland | 5 | S | N |
| | ITE | 10 | UK | N |
| Geochemistry | AGRG | 5 | E + W | N |
| Climate | ITE | 10 | UK | N |
| Water features | ITE | 10 | UK | N |
| Vegetation | ITE | 10 | GB | N |
| Fauna and flora | ITE | 10 | UK | N |
| Agriculture | MAFF | 5 | E + W | N |
| | DAFS | 5 | Scot. | N |
| | MANI | 10 | N. Ire. | N |
| | ITE | 10 | GB | N |
| Land quality | ITE | 10 | GB | N |
| Pollution | various | varies | GB | N |
| Population | OPCS/Birkbeck | 1 | GB | N |

*Key:* OSGB: Ordnance Survey of Great Britain; OSNI: Ordnance Survey of Northern Ireland; BGS: British Geological Survey; ITE: Institute of Terrestrial Ecology; AGRG: Applied Geochemistry Research Group, Imperial College; MAFF: Ministry of Agriculture, Fisheries and Foods; DAFS: Department of Agriculture and Fisheries for Scotland; MANI: Ministry of Agriculture for Northern Ireland; C: Community disk; N: National disk; OPCS: Office of Population Censuses and Surveys.

*Source:* Rhind and Mounsey (1986)

mapped at this detail) than are the original data. In addition, the requirement to have clear video images of background maps to underlay statistical or thematic maps ensured that some 1500 simple monochrome plots had to be generated from OS digital 1/625 000 scale map data, photographed and then indexed so that the 'best fit' background map is selected automatically in response to user nomination of an area of interest.

Though the initial way into the contents of the National Disk is via the Domesday gallery (Plate VII(c)), most users soon turn to the use of a hierarchical thesaurus. From the uppermost level of four topics (Culture, Economy, Environment and Society), a 7-level cross-linked structure expands to give over 9000 basic terms by which access is gained to text, pictures or digital data (Lee, 1987). Alternatively, keywords (such as 'Population Density' or 'Cretaceous rocks') may be specified; the system will then search the index and present the user with a list of all successful full or partial matches, from which a further selection may be made.

*Functionality* The capabilities of the system so far as spatial data are concerned can be defined formally as the ability to:

(i) view topographic maps, to run through the entire country and to 'zoom in' from national to very local views by moving from small-scale to large-scale maps of the same area;

(ii) measure area and distance in metric or imperial units by indicating boundaries or routes on the video topographic maps; measure National Grid references to 100-metre resolution by pointing at a map;
(iii) view satellite imagery, air photographs, slides and text as well as maps; to store relationships between these entities and retrieve in accordance with the relationships;
(iv) retrieve data by area name, by coordinate position, by pre-defined geographical 'window' and/or by variable;
(v) plot digital spatial data with either default or user-specified class intervals, select a desired colour scheme, areal limits, data resolution, etc.;
(vi) overlay these plots, if desired, on background topographic maps;
(vii) interrogate the display by pointing a cursor on the screen to obtain the value (or area name) at a point of interest;
(viii) compute the statistical correspondence between selected variables in the geographical 'window' selected;
(ix) dump selected data from the video disk onto floppy disk and also incorporate on the screen the user's own data supplied on floppy disk, together with that from the LV ROM;
(x) leave an 'audit trail' where the user has 'been' in the system so that bookmarks may be created to guide others directly to items of interest. This is particularly valuable in a teaching environment.

Additional capabilities which are more relevant with the other data sets include the ability to display time series data (e.g. newspaper sales by region, by time) as moving images. In addition, nine surrogate walks are available. On these, the user 'walks' around a farm, a town, a house, etc., progressing and turning around at will and zooming in to examine items of particular interest (Plate VII(d)). This is achieved by holding many hundreds of photographs, taken at a series of defined points and stored as a linked sequence, which the user may navigate at will; a plan is available to show feasible routes and the current position of the observer.

*Domesday and the future*

It is clear from the above description that the initial Domesday system is indeed revolutionary. It has already stimulated great interest and academic use. In Gateshead, for instance, it led sixth formers to create their own, expanded Domesday multicolour book of the town to demonstrate its virtues to businessmen (Anon., 1987). The report of the government Committee of Enquiry on the Handling of Geographic Information, chaired by Lord Chorley (HMSO, 1987) singled out the Domesday system as a unique development and emphasized its role as a teaching tool. Nonetheless, the present system still has some shortcomings. The most serious of these are:

- the limited analytical capability of the present software. This stems from two factors; the time available to write reliable basic, let alone sophisticated, spatial analytic software and the memory constraint imposed by the use of the BBC micro-computer. The latter also limits the number of colours which may be shown on the screen at any one time. Clearly, the advent of an IBM PC and other versions of the machine will reduce both problems. In the medium term, the move from a now obsolescent 8 bit processor to 32 bit processors will transform the analytical capabilities and will further extend the graphics interface, provide split screen working and other contemporary features.
- the lack of any regular updates for those data which are needed in highly topical form. Plans have already been laid for regular updating and publishing of new video disks, subject to sufficient user demand.
- the limitation of the data base to Britain. Over a dozen other countries have already expressed strong interest in replicating the Domesday project; within Britain, compilation of a new rural heritage disk is about to begin and several others (including one on London) are planned to follow.
- the possible misuse of data in combination through analyses carried out by unskilled users. It is evident that certain combinations (e.g. altitude with per cent unemployed)

are probably meaningless; more seriously, combining data derived from, say, maps at widely different scales may be most misleading. In the longer term, the only solution to this problem—encountered by all GIS—is to install a suitable Expert System front-end processor [see R and D recommendations in the Chorley Report (HMSO, 1987)]. In the short term, however, human guidance and education is the only solution.

The Domesday Project has already had a major impact in the way in which it has removed certain data from the private domain of data gatherers into the public domain. In some cases, data gatherers have been loath to lose control over their data, notably where these data ensure a steady cash return for copyright revenues. Indeed, the whole issue of copyright in a UK context is one which is threatened generally by developments in computing. Despite recent amendments to the Copyright Act which clarified the situation by bringing digital data clearly within the terms of the legislation, enforcement is essentially impossible in a stand-alone and widely distributed computer system such as Domesday. Traditional concepts of assessing copyright fees hinged upon the number of copies made of a paper-based product; in the situation where data may be used to produce an infinite number of different displays and these appear only momentarily upon a computer screen, assessment of the level of use of individual data sets is impossible. Thus, assuming that there will be no move towards the abolition of copyright charges through adoption of the American concept of material being 'in the public domain', there is no recourse other than to make a 'best guess' basis for setting charges predicated upon what the market will bear. For Domesday, the copyright arrangements were complex but were completed on the basis of an understanding that all parties were engaged in a pioneering effort. It was accepted that small subsets of the data on the national disk could be printed off or copied to a floppy disk for use elsewhere. At the time of making copyright agreements with the Ordnance Survey for use of its maps, no simple means existed for making 'hard copy' on paper of the television images; subsequently, new technology has rendered this relatively straightforward but the likely impact on OS paper map sales still seems certain to be small. How well these arrangements work in practice may affect the willingness of organizations and government agencies to repeat their commitment to Domesday. For them, it offers the opportunity of wider use of and publicity for the data resulting from their own efforts, allied in some cases to a small royalty payment. Whether this clashes with the objectives of the government's Tradeable Information Initiative—intended to sell off at the best possible price all marketable knowledge held by central government—remains to be seen.

Despite the present shortcomings, we believe the Domesday machine to be of critical importance for three reasons. The first of these stems from its ability to function with analogue as well as digital data; it is totally unreasonable to expect all historical and contemporary data to be converted into computer form before use. The second reason is that, because of its creation by an information-oriented organization rather than by geographers or surveyors, it treates GIS capabilities as just one, though an important one, set of data base operations; as a consequence, the long-fostered artificial separation between GIS-type data and other data is demonstrated to be chimerical and is avoided. Finally, because of these two factors and because of its exceptionally low price, Domesday and its successors may well become the most widely used information system in the world. As such, it will make much information equally available to the government planner, the commercial sector developer and the members of the lay public; indeed it could well have an important and positive role in the information society. That much of this information is geographical in nature will ensure that many more 'amateur' geographical practitioners come to exist. This has obvious implications for training needs. It also underlines the importance of existing professional organizations extending their view of both clients and members.

*Acknowledgements*

The authors were only members of a large collection of individuals working on this project (though Peter Armstrong was in overall charge). We readily acknowledge that most of what is reported here was a team achievement and thank all of our colleagues

for the high level of collaboration achieved in often trying circumstances. In particular, we wish to thank all the geographers—in schools, colleges, universities and elsewhere—who helped to ensure that the data for Domesday could be made available.

The colour illustrations were financed by grants kindly made available by the British Academy and by Birkbeck College, University of London.

*References*

Anon. 1987 *The Gateshead Domesday Book*, Gateshead Metropolitan Borough Council.

Goddard, J. B. and Armstrong, P. 1986 The 1986 Domesday Project. *Trans. Inst. Brit. Geogr.* NS **11**, 3: 290–5.

HMSO. 1987 *Handling Geographic Information.* The report of the Committee of Enquiry chaired by Lord Chorly. London: HMSO.

Lee, D. 1987 Indexing the Domesday Project. *The Indexer* **15,** 3: 145–50.

Openshaw, S., Wymer, C. and Charlton, M. 1986 A geographical information and mapping system for the BBC Domesday optical discs. *Trans. Inst. Brit. Geogr.* NS **11,** 3: 296–304.

Owen, D. W., Green, A. E. and Coombes, M. G. 1986 Using the social and economic data on the BBC Domesday disc. *Trans. Inst. Brit. Geogr.* NS **11,** 3: 305–14.

Rhind, D. W. and Mounsey, H. M. 1986 The land and people of Britain: a Domesday record. *Trans. Inst. Brit. Geogr.* NS **11,** 3: 315–26.

Wells, M. 1984 The JANET project. *University Computing* **6:** 56–62.